MA547 **Spring 2013**

PROJECTS **EMPIRICAL INVESTIGATIONS**

Assigned: 4/2/13 Due: 4/23/13

This project is worth 15% of the course. Please form a team with three-four students. You need a team with one member who can compute, one member who can do some statistical methods and one member who can write well. Your team members should decide what are the most valuable activities to do in the project. Each team should select one of the projects and only one team can do a specific project. A project report is required.

P #1. For a simple random sample of size n from a finite population of size N, an approximate 95% confidence interval for the finite population mean is $\bar{y} \pm 1.96\sqrt{(1-f)s^2/n}$. Draw populations of size N from Normal(μ, σ^2) and samples of size n to investigate the coverage properties of the confidence interval.

P #2. For a simple random sample of size n from a finite population of size N, an approximate 95% confidence interval for the coefficient of variation of the population mean is $\hat{cv} \pm 1.96\hat{cv}^2$. Draw populations of size N from Normal(μ, σ^2) and samples of size n to investigate the coverage properties of the confidence interval.

P #3. A simple random sample of size n is taken from a finite population of size N and two variables x and y are taken. Suppose the finite population mean of Y is needed and the finite population mean of X is known. A sensible estimator of $\bar{Y}$ is $\hat{\bar{y}} = \frac{\bar{y}}{\bar{x}}\bar{X}$. Compare this estimator with $\bar{y}$ and investigate its superiority by considering bias and mean squared error using a simulation study.

P #4. Consider various numbers S of strata with respect to the Mantel Haenszel test, use Monte Carlo methods to find the P-value of the test

of no treatment effect. You are to select data from S hypergeometric distributions.

P #5. Consider the McNemar's test to assess the effect an exercise program. Suppose the study is on blood pressure and there are 1000 subjects and measurements are taken when the program starts and ends. In such a study age is an important variable. Simulate data of this form. Then, poststratify the data using a logistic regression procedure to say, 5 strata. Finally, for each stratum obtain the P-value of the test. Also obtain a combined P-value for all strata

P #6. Consider the mercury in blood data on methylmercury in fish and human chromosome damage. Obtain the P-value of a test of no treatment effect using a *bootstrap* method. Bootstrapping is done as follows. Draw two samples of size 16 and 23 with replacement from the control and treatment groups. Compute the test statistic and repeat the process, say 10,000 times. Use two sets of draws, one set with equal probabilities and and the other with unequal probabilities and compare your answers.

P #7. Consider the data on micronuclei harboring whole chromosomes in alcoholics. Obtain the P-value of a test of no treatment effect using *bootstrap* method. Bootstrapping is done as follows. Draw a sample of size 20 with replacement from the 20 subjects. Compute the test statistic and repeat the process, say 10,000 times. Use two sets of draws, one set with equal probabilities and and the other with unequal probabilities and compare your answers.